What is claimed is:

1. A method for organizing a plurality of data files using meta data, having at least one meta data element, at least associated with each data file, the method comprising:

extracting, for at least some of the data files, at least one meta-data element associated with that data file;

organizing the extracted meta-data elements in a desired order based on values for the extracted meta-data elements;

inputting at least one parameter value; and

dividing at least some of the data files into groups based on the extracted meta-data elements and the input parameter value.

2. The method of claim 1, wherein dividing the at least some data files comprises determining, for each of at least one of the at least one parameter value, a similarity value for at least two of the plurality of data files using at least some of the extracted meta-data elements and that parameter value.

3. The method of claim 2, wherein determining the at least one similarity value comprises determining the at least one similarity value as:

$$S_K(i,j) = \exp\left(-\frac{|t_i - t_j|}{K}\right),$$

where:

$S_K(i,j)$ is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

K is the parameter value; and

$t_i$ and $t_j$ are actual values of at least one meta-data element of the at least one extracted meta-data elements for the $i^{th}$ and $j^{th}$ data files.

4. The method of claim 2, wherein determining the at least one similarity value comprises determining the at least one similarity value as:

$$S_K(i,j) = \exp\left(\frac{1}{K}\left(\frac{< v_i, v_j >}{|v_i||v_j|} - 1\right)\right) .$$

where:

$S_K(i,j)$ is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

K is the parameter value; and

$v_i$ and $v_j$ are actual vector values determined from the ith and the jth data files.

5. The method of claim 2, further comprising determining, for each of at least some data files, at least one novelty value for that data file based on the at least one similarity value for that data file and for a number of nearby data files.

6. The method of claim 5, wherein determining at least one novelty value comprises determining at least one novelty value as:

$$v_K(s) = \sum_{l,n=-5}^{5} S_K(s+l, s+n)\, g(l,n) \quad .$$

where:

$v_K(s)$ is the novelty value; and

g is a Gaussian tapered 11 x 11 checkerboard kernel.

7. The method of claim 5, further comprising determining at least one boundary location between ones of the plurality of data files based on the at least one novelty value determined for at least some of the data files.

8. The method of claim 7, further comprising determining, for at least some of the determined boundary locations, a confidence value for that boundary location.

9. The method of claim 8, wherein determining a confidence value for a boundary location comprises determining the confidence value as:

$$C(B_K) = \sum_{l=1}^{|B_K|-1} \frac{1}{(b_{l+1}-b_l)^2} \sum_{i,j=b_l}^{b_{l+1}} S_K(i,j) - \sum_{l=1}^{|B_K|-2} \frac{1}{(b_{l+1}-b_l)(b_{l+2}-b_{l+1})} \sum_{i=b_l}^{b_{l+1}} \sum_{j=b_{l+1}}^{b_{l+2}} S_K(i,j) \quad .$$

where:

$C(B_K)$ is the confidence value for the $B_K^{th}$ boundary;

$S_K(i,j)$ is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

b is the index value of detected boundary at a particular value for the input parameter K level.

10. The method of claim 8, further comprising determining, for at least some of the determined boundary locations, at least one of the at least one parameter value that maximizes the confidence value.

11. A method for organizing a plurality of data files using meta-data having at least one meta-data element that is at least associated with a corresponding one of the data files, the method comprising:

processing at least one set of meta-data, where each meta-data corresponds to a data file;

obtaining a desired value for analyzing the meta-data; and

determining a structure within the set of meta-data elements using an obtained parameter value, wherein the structure is determined by comparing, for at least a subset of the plurality of data files, at least a subset of the meta-data using the parameter value to each other.

12. The method of claim 11, further comprising clustering the data files into groups using the determined structure of the meta-data.

13. The method of claim 12, further comprising determining boundaries from the determined clusters of data files, wherein the boundaries are located between the determined clusters of data files.

14. The method of claim 13, further comprising:

determining a similarity value by comparing at least some of the meta-data elements in one cluster of data files to at least some other ones of the meta data elements in that element cluster of data files; and

determining a dissimilarity value by comparing at least some of the meta-data elements in one cluster of data files to at least some of the meta-data elements in another cluster of data files.

15. The method of claim 14, further comprising:

determining a value corresponding to a desired grouping of the clusters of data files based on the differences of the similarity values and the dissimilarity values.

16. A storage medium storing a set of program instructions executable on a data processing device and usable to organize a plurality of data files by using meta data having at least one meta data element at least associated with each data file, the program comprising:

instructions for extracting for at least some of the data files, at least one meta-data element associated with that data file;

instructions for organizing the extracted meta-data elements in a desired order based on values for the extracted meta-data elements;

instructions for inputting a parameter value; and

instructions for dividing at least some of the data files into groups based on the extracted meta-data elements and the input parameter value.

17. The storage medium of claim 16, instructions for dividing at least some of the data files into groups further comprising instructions for determining, for each of at least one of the at least one parameter value, a similarity value for at least two of the plurality of data files using at least some of the extracted meta-data elements and that parameter value.

18. The storage medium of claim 17, further comprising instructions for determining, for each of at least some data files, at least one novelty value for that data file based on the at least one similarity value for that data file and for a number of nearby data files.

19. The storage medium of claim 17, wherein instructions for determining the at least one similarity value comprises instructions for determining the at least one similarity value as:

$$S_K(i,j) = \exp\left(-\frac{|t_i - t_j|}{K}\right),$$

where:

$S_K(i,j)$ is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

K is the parameter value; and

$t_i$ and $t_j$ are actual values of at least one meta-data element of the at least one extracted meta-data element for the $i^{th}$ and $j^{th}$ data files.

20. The storage medium of claim 17, wherein instructions for determining the at least one similarity value comprises instructions for determining the at least one similarity value as:

$$S_K(i,j) = \exp\left(\frac{1}{K}\left(\frac{< v_i, v_j >}{|v_i||v_j|} - 1\right)\right) \quad .$$

where:

$S_K$ (i,j) is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

K is the parameter value; and

$v_i$ and $v_j$ that are actual vector values determined from the ith and the jth data files.

21. The storage medium of claim 18, further comprising instructions for determining at least one boundary location between ones of the plurality of data files based on the at least one novelty value determined for at least some of the data files.

22. The storage medium of claim 18, wherein instructions for determining at least one novelty value comprises instructions for determining the at least one novelty value as:

$$v_K(s) = \sum_{l,n=-5}^{5} S_K(s+l, s+n)\, g(l,n) \quad .$$

where:

$v_K(s)$ is the novelty value; and

g is the Gaussian tapered 11 x 11 checkerboard kernel.

23. The storage medium of claim 21, further comprising instructions for determining, for at least some of the determined boundary locations, a confidence value for that boundary location.

24. The storage medium of claim 23, wherein instructions for determining at least one confidence value comprises instructions for determining each of such confidence value as:

$$C(B_K) = \sum_{l=1}^{|B_K|-1} \frac{1}{(b_{l+1} - b_l)^2} \sum_{i,j=b_l}^{b_{l+1}} S_K(i,j) - \sum_{l=1}^{|B_K|-2} \frac{1}{(b_{l+1} - b_l)(b_{l+2} - b_{l+1})} \sum_{i=b_l}^{b_{l+1}} \sum_{j=b_{l+1}}^{b_{l+2}} S_K(i,j) \quad .$$

where:

$C(B_K)$ is the confidence value for the $B_K{}^{th}$ boundary;

$S_K$ (i,j) is the similarity value for the $i^{th}$ data file and the $j^{th}$ data file;

b is the detected boundary at a level.

25. The storage medium of claim 23, further comprising instructions for determining, for at least some of the determined boundary locations, at least one of the at least one parameter value that maximizes the confidence value.

26. A data file organizing system usable to organize a plurality of data files using meta data having at least one meta data element that is at least associated with a corresponding one of the data files, comprising:

a meta-data extracting circuit, routine, or application that extracts, for at least some of the data files, at least one meta-data element associated with that data file;

a meta-data organizing circuit, routine or application that organizes the extracted meta-data elements in a desired order based on values for the extracted meta-data elements;

a similarity value determining circuit, routine or application that determines, for at least one of the at least one parameter value, a similarity value for at least two of the plurality of data files using at least some of the extracted meta-data elements and that parameter value

a novelty value determining circuit, routine or application that determines at least one novelty value for that data file based on the at least one similarity value for that data file and for a number of nearby data files;

a data dividing determining circuit, routine or application that divides at least some of the data files into groups based on the extracted meta-data elements and the input parameter value by determining at least one boundary location between ones of the plurality of data files based on the at least one novelty value determined for at least some of the data files; and

a confidence value determining circuit, routine or application that determines, for at least some of the determined boundary locations, a confidence value for that boundary location, wherein the data dividing circuit, routine, or application further determines, for at least some of the determined boundary locations, the at least one parameter value that maximizes the confidence value.